Article

Integrative Modelling Coupled with Ion Mobility Mass Spectrometry Reveals Structural Features of the Clamp Loader in Complex with Single-Stranded DNA Binding Protein

Argyris Politis[1]†, Ah Young Park[1]†, Zoe Hall[1], Brandon T. Ruotolo[2] and Carol V. Robinson[1]

*1 - **Department of Chemistry,** University of Oxford, South Parks Road, Oxford OX1 3QZ, UK*
*2 - **Department of Chemistry,** University of Michigan, 930 North University Avenue, Ann Arbor, MI 48109, USA*

***Correspondence to Argyris Politis and Carol V. Robinson:** argyris.politis@chem.ox.ac.uk; carol.robinson@chem.ox.ac.uk*
http://dx.doi.org/10.1016/j.jmb.2013.04.006
Edited by M. Debatisse-Buttin

Abstract

DNA polymerase III, a decameric 420-kDa assembly, simultaneously replicates both strands of the chromosome in *Escherichia coli*. A subassembly of this holoenzyme, the seven-subunit clamp loader complex, is responsible for loading the sliding clamp (β_2) onto DNA. Here, we use structural information derived from ion mobility mass spectrometry (IM-MS) to build three-dimensional models of one form of the full clamp loader complex, $\gamma_3\delta\delta'\psi\chi$ (254 kDa). By probing the interaction between the clamp loader and a single-stranded DNA (ssDNA) binding protein (SSB_4) and by identifying two distinct conformational states, with and without ssDNA, we assemble models of $\psi\chi$–SSB_4 (108 kDa) and the clamp loader–SSB_4 (340 kDa) consistent with IM data. A significant increase in measured collision cross-section (~10%) of the clamp loader–SSB_4 complex upon DNA binding suggests large conformational rearrangements. This DNA bound conformation represents the active state and, along with the presence of $\psi\chi$, stabilises the clamp loader–SSB_4 complex. Overall, this study of a large heteromeric complex analysed by IM-MS, coupled with integrative modelling, highlights the potential of such an approach to reveal structural features of previously unknown complexes of high biological importance.

Introduction

The *Escherichia coli* DNA polymerase III replisome copies the chromosomal DNA in a rapid and highly processive manner, facilitated by the ring-shaped sliding clamp β_2.[1–4] A subassembly of the replisome, the seven-subunit clamp loader, opens and loads the β_2 ring onto DNA in an ATP-driven process.[5,6] The clamp loader comprises a central pentameric ring ($\gamma_3\delta\delta'$), which is attached to a $\psi\chi$ heterodimer via interaction of ψ with the three γ/τ subunits[7]; γ is a truncated version of τ produced by programmed translation frameshifting,[8] and the $\gamma_3\delta\delta'$ form is active in clamp loading but not in processive DNA synthesis.[9] The χ subunit mediates interaction between the clamp loader and the tetrameric single-stranded DNA (ssDNA) binding protein SSB_4.[6,10,11] This interaction, which occurs during replication, stabilises the complex and, at high ionic strength, stabilises polymerase–template–primer interactions.[4,12] The primary role of SSB_4 is to protect ssDNA from being degraded and to maintain cooperation with DNA binding proteins.[4,13–15] Upon DNA binding, a switch between active and inactive states of SSB_4 occurs, which further enhances the activity of the clamp loader.[16] Although X-ray crystal structures are available for $\gamma_3\delta\delta'$, $\psi\chi$ with the SSB-Ct peptide[17] and SSB_4, the latter with and without DNA,[7,18] complete structures for $\gamma_3\delta\delta'\psi\chi$, $\psi\chi$–SSB_4, and $\gamma_3\delta\delta'\psi\chi$–$SSB_4$–(ssDNA) assemblies have not been reported.

Over the last few years, mass spectrometry (MS) has emerged as a powerful tool for integrative structural biology. MS is now being used to construct topological maps of large, multimeric protein complexes.[19–21]

Where high-resolution data are available for some components, structure prediction can be made with atomic resolution.[22,23] Coupling of ion mobility (IM) to MS, in which the collision cross-section (CCS) of a protein complex can be deduced, has proved highly informative for the study of the structure and dynamics of macromolecules.[24,25] Within the last decade, the development of IM-MS approaches, enabling the study of heterogeneous and dynamic protein assemblies, has provided structural insight into those biological targets that are the most challenging for structural biology.[26–30] Recently, IM-MS has been employed to distinguish between topologies for two subcomplexes in the eukaryotic initiation factor 3,[31] to establish the stability and topology of two tetrameric membrane proteins,[32] and to elucidate the subunit connectivity and stability of two CRISPR-associated proteins, the Cascade and Csy complexes.[33] Whilst IM-MS has been used to propose structural models,[34,35] it has only recently been assessed more quantitatively as a tool for structure prediction.[36] The use of experimentally determined CCS as restraints for predicting the topology of heteromeric assemblies was explored, revealing good agreement between the best-scored models and their corresponding X-ray crystal structures.[36]

In this study, we use restraints from IM-MS to build structural models of assembly intermediates observed in our experiments, as well as the clamp loader in complex with SSB_4 (Fig. 1). Model structures for the minimal five-subunit clamp loader ($\gamma_3\delta\delta'$) and the tetrameric SSB_4 have been reported previously from IM-MS and crystallographic data.[37] Here, we extend our initial studies and use MS to define multiple subcomplexes and to construct an assembly pathway of the full clamp loader complex ($\gamma_3\delta\delta'\psi\chi$) bound to SSB_4, in the presence or absence of ssDNA. The connectivity and stoichiometry determined by these experiments, together with CCSs from IM, are used as restraints to build models of the intermediates formed along the pathway. In this process, we begin by assembling a model for $\gamma_3\delta\delta'\psi\chi$ and then probe the interaction between the $\psi\chi$ heterodimer and SSB_4. Together, this information enables us to build an accurate model of $\psi\chi$–SSB_4. Finally, we investigate structural features of the eleven-subunit clamp loader bound to SSB_4 and conformational changes induced upon binding of ssDNA.

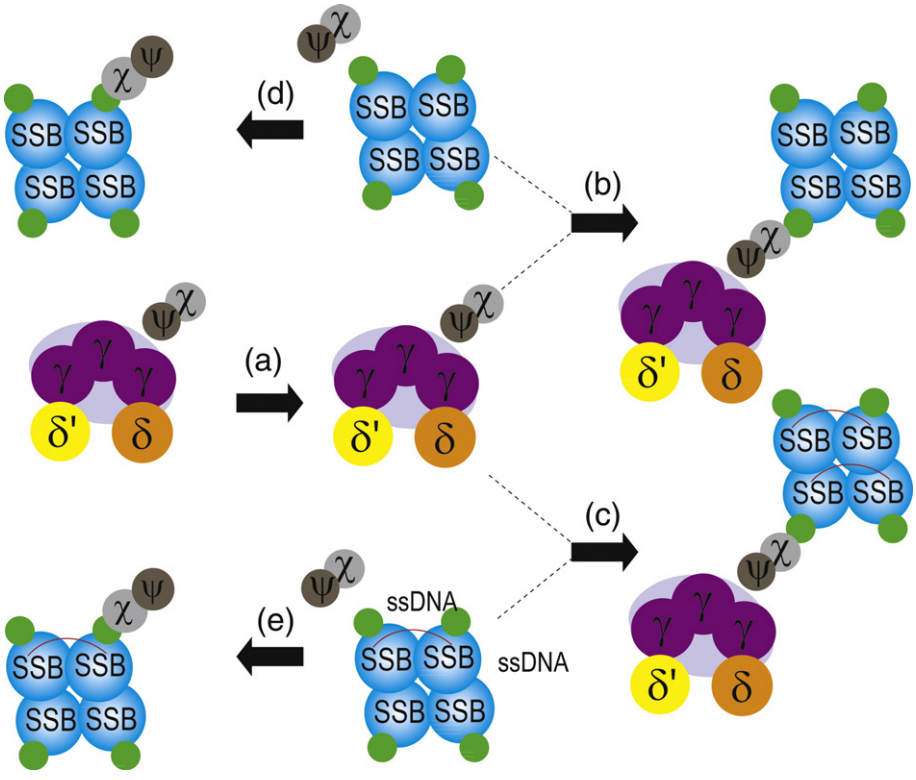

Fig. 1. Assembly pathway of the clamp loader–SSB_4 complex and its intermediate subcomplexes (a). The minimal clamp loader ($\gamma_3\delta\delta'$) interacts with $\psi\chi$ heterodimer to form the complete clamp loader complex $\gamma_3\delta\delta'\psi\chi$ (b). The clamp loader binds to tetrameric SSB to form the $\gamma_3\delta\delta'\psi\chi$–$SSB_4$ complex (c). The conformational change of SSB_4 upon binding of ssDNA does not affect the interaction with the clamp loader, which forms the $\gamma_3\delta\delta'\psi\chi$–$SSB_4$–ssDNA complex. $\psi\chi$ also interacts with SSB_4 to form the subassemblies $\psi\chi$–SSB_4 (d) and $\psi\chi$–SSB_4–ssDNA (e). The ssDNA is depicted by the red curved lines on the SSB protein. The N-terminal domains of SSB, shown in green, indicate the proposed binding site to $\psi\chi$.

Results

Modelling approach

To apply the information derived from IM-MS of protein assemblies and their subcomplexes to structure characterisation, we use a coarse-grained (CG) modelling approach as described in Fig. 2. Briefly, proteins or protein domains are represented as spheres, and protein complexes and their subcomplexes are represented as a set of overlapping spheres, with radius r defined using the subunit (or domain) CCS, $r = (\Omega/\pi)^{1/2} - r_{He}$, where r_{He} is the correction of the drift gas (radius of helium) contribution to the CCS. Models for protein complexes are built in a stepwise manner by combining subunits and subcomplexes, which make up the assembly. To account for flexibility, we subjected the initial atomic coordinates to dynamical analysis using NAMD (Fig. S1).[39] Thus, the atomic coordinates of the subcomplexes studied here were refined by performing energy minimisation (see Supplementary Data). Such analysis is performed for all subcomplexes used to build the assembly and allowed us to eliminate any steric clashes in the final models and to search for an energetically favourable conformations, which used as starting structures. Adequate sampling, a critical step of this modelling approach, was obtained by applying a Monte Carlo search followed by a conjugate gradient optimisation step. All degrees of freedom are considered, rendering this approach unbiased towards any particular solution. Candidate models are evaluated by incorporating a scoring function that encodes restraints derived from IM-MS data. The best-scoring models are clustered based on their structural similarities using a hierarchical tree approach.

The seven-subunit clamp loader complex $\gamma_3\delta\delta'\psi\chi$

Here, we set out to assemble a structural model for the seven-subunit clamp loader complex using restraints from IM-MS. The stoichiometry and homogeneity of the subcomplexes, including $\psi\chi$, $\gamma_3\delta\delta'$, and $\gamma_3\delta\delta'\psi\chi$, were determined previously[7,9,37,40]; however, their overall topological arrangement was not examined. To study the topology of the clamp loader complex, we begun by recording the spectra for both $\gamma_3\delta\delta'$ and $\gamma_3\delta\delta'\psi\chi$ where we observed only one charge state series that corresponds to each

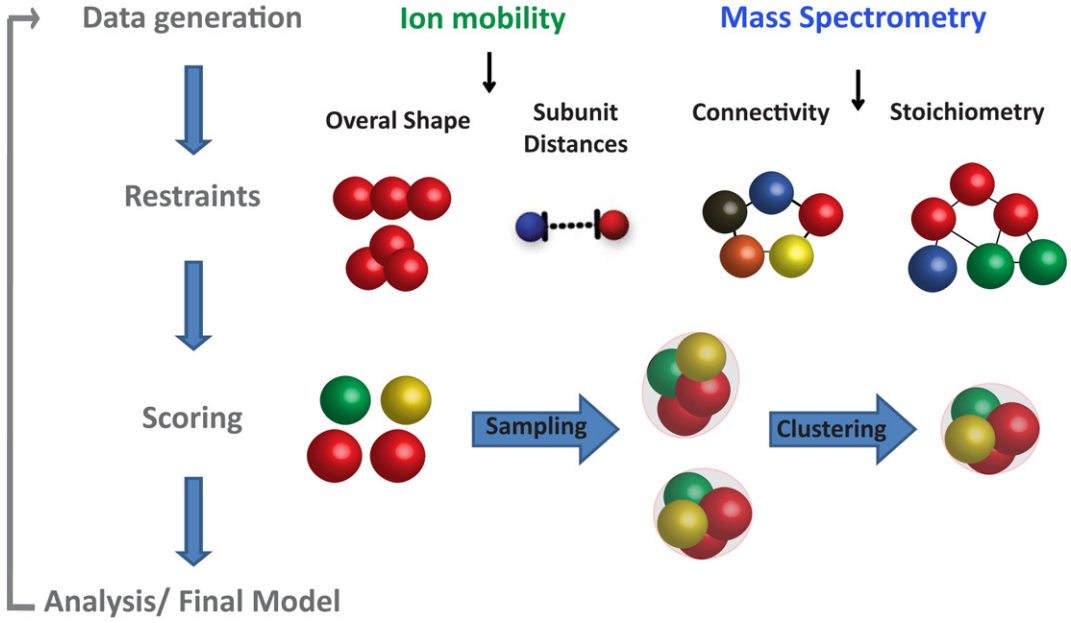

Fig. 2. Schematic overview of the modelling approach. The steps taken are as follows: Data generation and interpretation: IM-MS data are recorded, including information about subunit connectivity, stoichiometry, and the overall shape of protein assemblies. Based on these data, topological maps can be constructed.[20,22] Both atomic- and low-resolution CG representations can be used for modelling. Spatial restraints: Experimental IM-MS values are encoded into restraints that are incorporated into a single scoring function for evaluating structural models. Such restraints examine the inter-subunit interactions (overlap and connectivity restraint) and the overall shape of the assemblies (CCS and volume restraints). Models are generated by fully sampling the conformational space using a Monte Carlo conjugate gradient search as implemented in the IMP (*I*ntegrative *M*odelling *P*latform).[38] Scoring: Model structures are scored using a harmonic function that measures the deviation of the calculated model CCS from the experimental CCS. Analysis: The ensemble of best-scored structures is clustered and the representative models are further refined to provide the final structures.

protein complex centred at 31+ and 32+, respectively (Fig. 3a and b). Analysis of the peak shape and width of the arrival time distributions (ATDs) from the IM experiments, expressed as $t/\Delta t$ (centroid values of ATD peaks normalised to their widths at half-height), indicated a single conformation or a conformational family of closely related structures ($t/\Delta t$, ranges from 10 to 15). The ATD of the ions (averaged for all charge states) was then converted to yield orientationally averaged CCSs and used as a restraint for modelling analysis.

To model the $\gamma_3\delta\delta'\psi\chi$ complex, we used a CG representation scheme, whereby each domain was represented as a sphere, scaled by the calculated CCS for the corresponding atomic coordinates.[36,37] Missing residues were accounted for by adding an additional sphere for each γ subunit (green spheres; Fig. 3c and d) scaled by their mass.[36,37] The degree of overlap between spheres was determined using the corresponding X-ray coordinates. Where no atomic information for subunit interfaces was available, an average value (29 ± 10% with an additional tolerance of 5%) was used to guide our search for model structures.[36] To assemble the clamp loader complex, $\gamma_3\delta\delta'$-$\psi\chi$, we begin by generating CG models for its building blocks, namely, $\gamma_3\delta\delta'$ and $\psi\chi$. For the minimal clamp loader ($\gamma_3\delta\delta'$), we make use of a model developed previously,[37] based on published crystal structures.[18] The CCS calculated for this complex (8936 Å^2) is in close agreement with the experimentally determined value (8904 Å^2), thus suggesting that no significant rearrangement occurred in the gas phase (Table 1). For $\psi\chi$, a model, which includes the missing residues, was built using CCS from IM and the X-ray crystal structure [Protein Data Bank (PDB) entry: 1EM8].[7] In this model, the χ

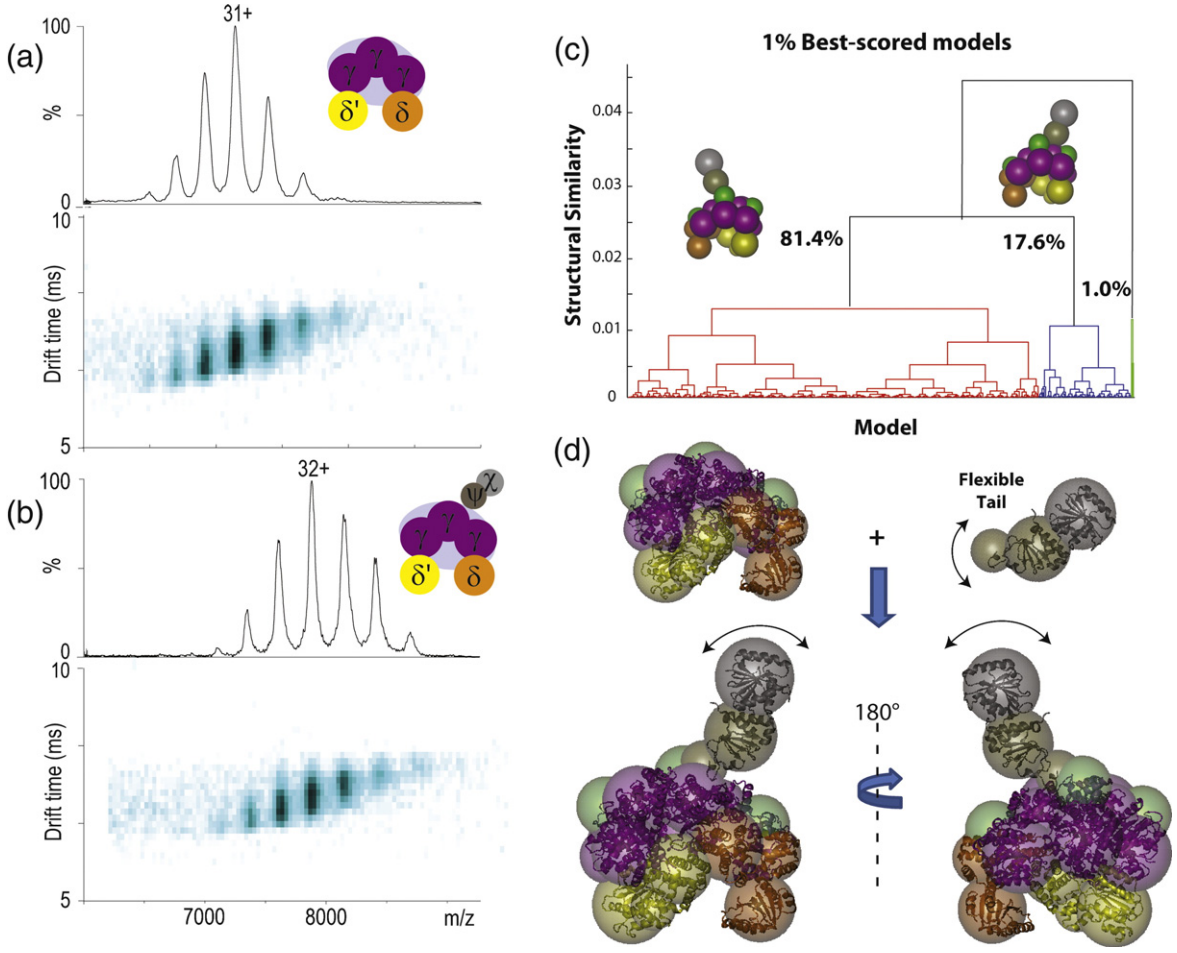

Fig. 3. IM-MS data and modelling analysis of the clamp loader complex. The mass spectra and IM contour plots superimposed on an *m/z* scale are shown for the clamp loaders (a) $\gamma_3\delta\delta'$ and (b) $\gamma_3\delta\delta'\psi\chi$. (c) Hierarchical clustering of 1% best-scored models from pairwise structural similarity scores reveals two major clusters representing 81.4% and 17.6%. The representative structure of the major cluster is selected as the final solution shown in (d). The best-scored model is determined after scoring and clustering a large number of structures. The model structure is shown in cartoon representation positioned by superimposing the optimised atomic coordinates onto the CG model structures. The missing residues are represented by green spheres scaled according to their mass.[37]

Table 1. Measured CCSs and masses of the protein subcomplexes studied here

Complexes	CCS[a] (Å^2)	Mass (kDa)	Figure	Structure
ψχ	2520	32.3	3	1EM8
$\gamma_3\delta\delta'$	8904	221.5	3c	1JR3
$\gamma_3\delta\delta'\psi\chi$	9434	253.8	3c	Model
SSB_4	4062	76.0	S1	1EQQ
SSB_4–ψχ	5708	108.3	4a	Model
SSB_4–ssDNA	4424	85.8	S1	1EYG
SSB_4 -ψχ–ssDNA	5897	117.7	4b	Model
SSB_4– $\gamma_3\delta\delta'\psi\chi$	11,529	337.0	5a	Model
SSB_4– $\gamma_3\delta\delta'\psi\chi$–$ssDNA_2$	12,488	340.6	5b	Model

[a] The experimental error is estimated to be ±3%.

subunit is represented as a single sphere, whereas ψ is represented as two overlapping spheres due to its elongated structure (Fig. 3c and d and Fig. S1a). The flexible N-terminus is represented by a sphere (Fig. S1b) where its distance from the other ψ sphere was adjusted from the CCS value of ψχ (Table 1).

Having assembled the CG models for the two individual subcomplexes, $\gamma_3\delta\delta'$ and ψχ, we turned our focus to modelling the full $\gamma_3\delta\delta'\psi\chi$ complex. We begin by sampling the conformational space for the seven-subunit complex using a Monte Carlo search.[36] Overall, 50,000 models were generated randomly by fixing the relative position of the subunits within the individual $\gamma_3\delta\delta'$ and ψχ subcomplexes and by enforcing connectivity between the N-terminus of ψ subunit and the γ subunits. This is in line with proposals that ψχ and the clamp loader are bridged by the interaction between the N-terminal segment of ψ subunit (small tan sphere) and the C-terminus of γ.[7] The degree to which the spheres overlap can be calculated by the centre of mass distances between spheres. In order to ensure physical interactions between subunits within the modelled structures, we applied an overlap filter. This eliminates any structures in which the spherical overlap is outside the range (15–45%) (Supplementary Data).[36] Following this, a total of 26,046 structures remained.

The next step was to score the remaining models using the volume and CCS restraints (see Supplementary Data). Both volume and CCS restraints are implemented as a harmonic penalty function that also takes into account the experimental error (2 × SDs = 6%, where SD is the standard deviation from the mean and is estimated to be 3%).[36] Therefore, all candidate models are scored based on their closeness to the experimental data and the best-scored models are selected for further analysis. Since multiple solutions are in agreement with the input restraints, we performed clustering of the best-scored models in order to investigate their structural variability in the ensemble of "good-scoring" models. The 1% best-scoring models (260 structures) were selected using a hierarchical clustering approach[43] (Supplementary Data). The pairwise scores between the structures are calculated using the ultrafast shape recognition (USR) score.[36] Briefly, USR has been used as an algorithm to compare molecular shapes.[44] Here, it is used to compare two CG models and the scores given range from 0 (non-related) to 1 (identical). For $\gamma_3\delta\delta'\psi\chi$, two main clusters were identified using a cutoff difference of USR score of 0.03 (Fig. 3a). The major cluster (81.4%) corresponds to a geometrical arrangement where ψχ extends over the two γ subunits adjacent to δ (Fig. 3c). The minor cluster, representing 17.6% of the top-scored models, reflects a more compact arrangement where the ψχ subcomplex points in the opposite direction towards the δ′ subunit. Each cluster is represented by the model with the best score within each cluster.

Binding ψχ to SSB_4 and the effects of ssDNA

Next, we investigated the architecture of the tetrameric SSB protein bound to χψ, with or without ssDNA. To determine the binding stoichiometry of SSB to ψχ, we incubated SSB_4 and ψχ in a 1:1 ratio and subjected it to analysis by MS. The MS spectrum revealed two charge state series, centred on the 16+ and 19+ charge states and corresponding to SSB_4 and SSB_4–ψχ (Fig. 4a). The peak intensities of SSB_4 relative to SSB_4–ψχ indicate that SSB_4 and ψχ form a weak interaction, under the experimental conditions employed. Similar results and binding stoichiometries were obtained at higher molar ratios of ψχ to SSB.

It has been proposed that, in the presence of ssDNA, the SSB binding affinity to χ increases by up to 1000-fold.[12,45] Therefore, the abovementioned experiment was repeated in the presence of ssDNA, in order to establish whether there is any increase in binding of χψ to SSB in its DNA bound form (Fig. 4b). Two charge state series were observed in the MS spectrum as major and minor species corresponding to SSB_4–ssDNA and ψχ–SSB_4 bound to only one ssDNA (d35-mer), respectively. Interestingly, by mixing SSB_4 with d65-mer and d35-mer in 1:2 and 1:3 ratio, respectively, we observed two ssDNAs bound on the tetramer (Fig. S4). Although the overall quality of the MS was impaired due to challenges experienced in obtaining a stable electrospray, an

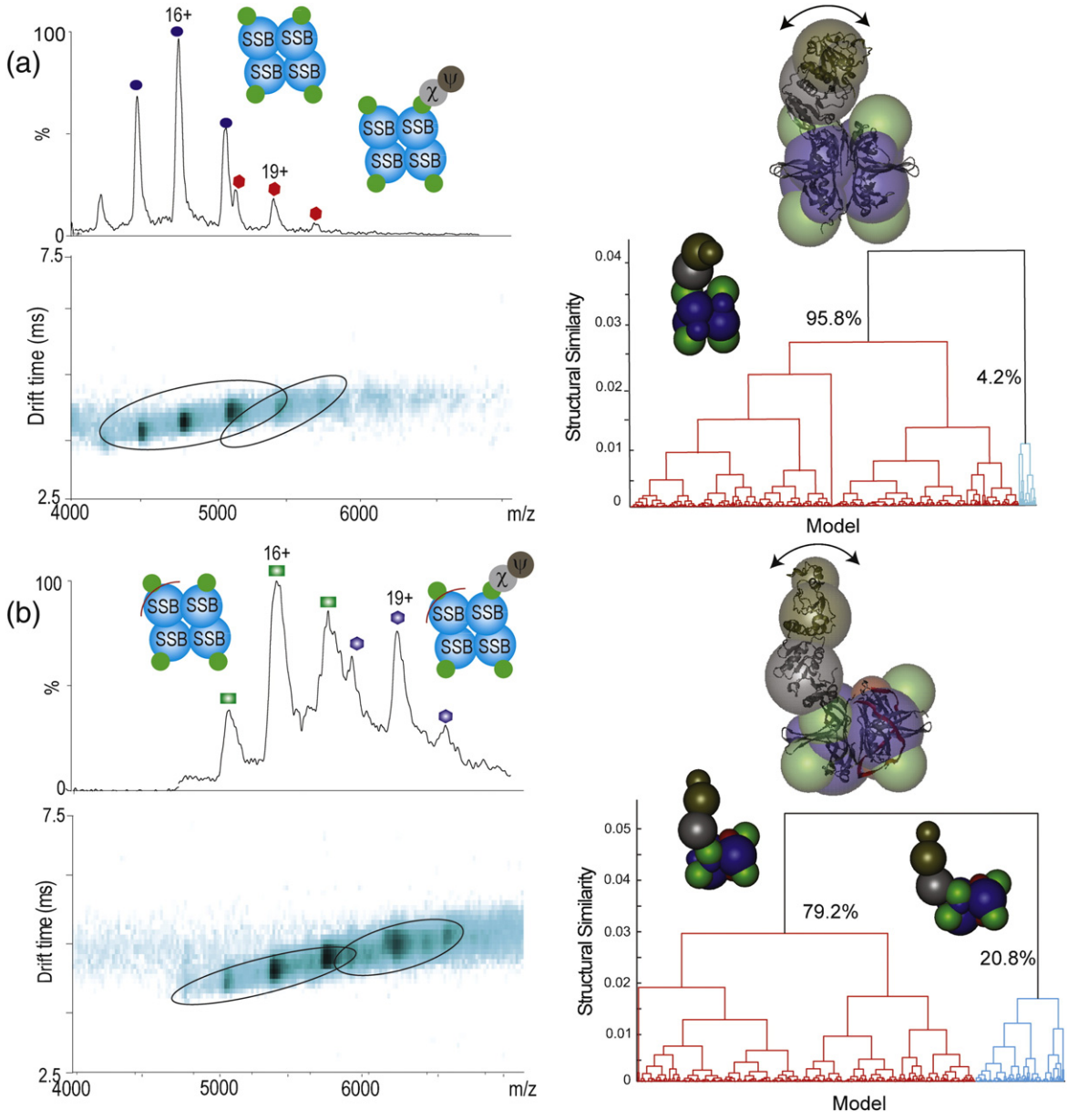

Fig. 4. IM-MS data and modelling analysis of ψχ–SSB_4 subassembly with and without ssDNA. The mass spectra and IM contour plots superimposed on an *m/z* scale for (a) ψχ–SSB_4 and (b) ψχ–SSB_4–ssDNA. A model for SSB_4 based on X-ray crystallography (PDB entry: 1EQQ[41]) and IM data is used together with a model for ψχ as input structures for generating candidate models. Scoring of model structures and subsequent clustering of the 1% best-scoring models reveals a major cluster (95.8%). The best-scored model within this cluster is chosen as the representative structure. The modelling strategy applied to the ψχ–SSB_4 subcomplex bound to ssDNA revealed conformational changes induced by ssDNA binding. The X-ray crystal structure of SSB_4 in complex with ssDNA was used as a starting point for modelling (PDB entry: 1EYG[42]). Clustering reveals two major clusters (79.2% and 20.8%). A representative structure of the major cluster is selected as the most likely solution for the ψχ–SSB_4–ssDNA subcomplex structure. Missing residues in SSB are represented by green spheres.

increase in the relative intensity of the peaks corresponding to ψχ–SSB_4–ssDNA over those observed for ψχ–SSB_4 was clearly observed. This supports the hypothesis that ssDNA assists in SSB_4 binding to χψ.

CCS was measured for SSB_4 and SSB_4–χψ using IM-MS, both with and without ssDNA (Table 1), and used as restraints in our modelling approach. From these data, we propose a model for SSB_4–χψ by assembling models for the constituent

subcomplexes SSB_4 and χψ. A structural model for SSB_4 was proposed previously by combining crystallographic information (PDB ID: 1EQQ)[41] and CG modelling[37] (Fig. S2a). In that model, each subunit was represented by three domains and the overall structure conformed to a symmetrical arrangement.

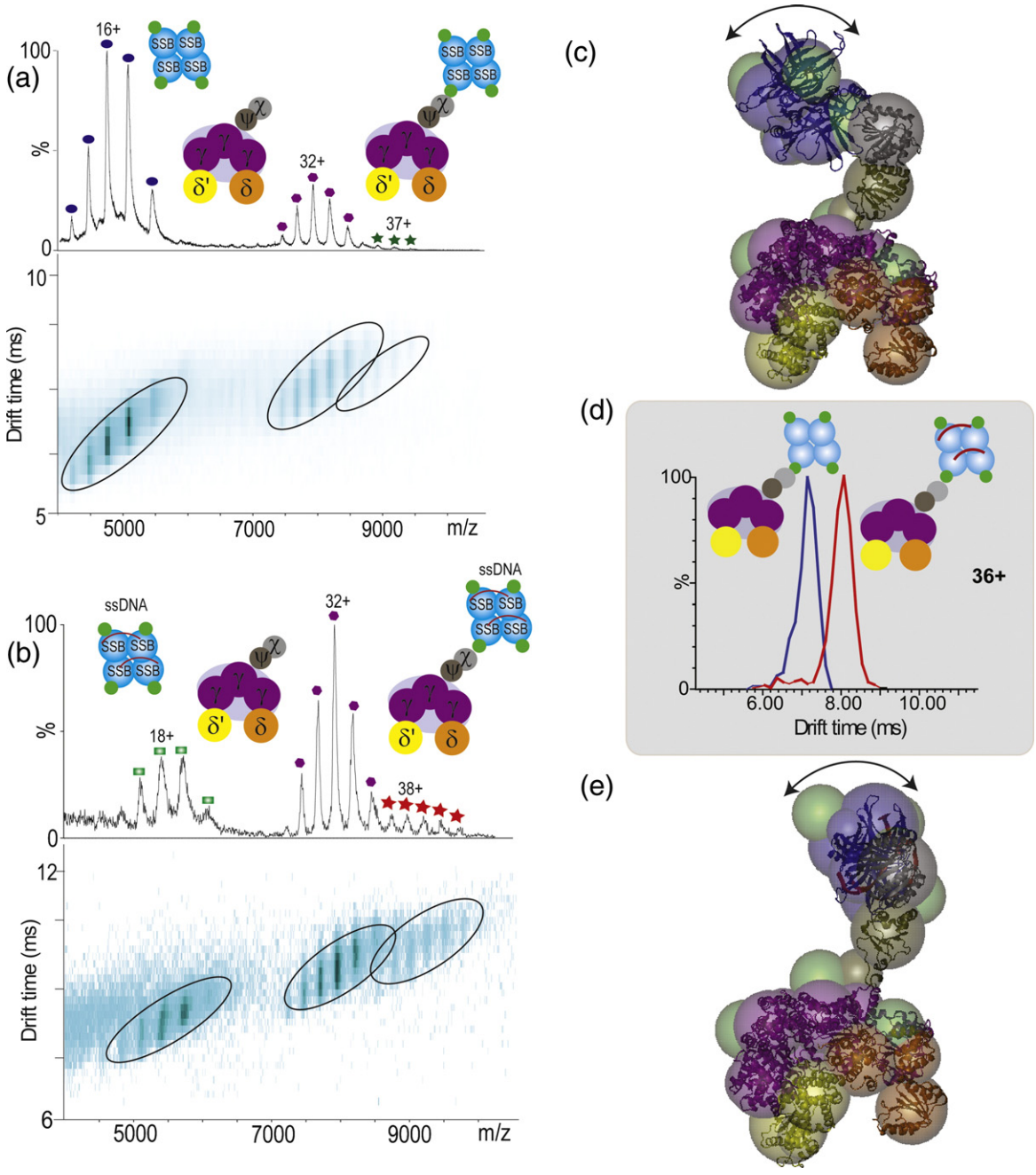

Fig. 5. Complete three-dimensional model of the clamp loader–SSB_4 complex in the absence and presence of ssDNA. The mass spectra and IM contour plots superimposed on an *m*/*z* scale are shown for (a) $\gamma_3\delta\delta'\psi\chi$–$SSB_4$ and (b) $\gamma_3\delta\delta'\psi\chi$–$SSB_4$–$ssDNA_2$. Complexes are shown above their corresponding charge state series. (c) A structural model of the $\gamma_3\delta\delta'$–ψχ–SSB_4 complex generated by integrating IM-MS restraints with molecular modelling, which uses a single scoring function. The intermediate sucbcomplexes of this complex, namely, $\gamma_3\delta\delta'$–ψχ and SSB_4, were used as input structures for sampling the conformational space. After implementation of restraints and further analysis, a structural model that shows the closest fit to the experimental data is suggested (d). Arrival time contours are overlaid for the 36+ charge state peaks for $\gamma_3\delta\delta'\psi\chi$–$SSB_4$ (blue) and $\gamma_3\delta\delta'\psi\chi$–$SSB_4$–$ssDNA_2$ (red) (e). The same strategy was followed to build a model structure for the clamp loader complex bound onto ssDNA. In such models, the conformational changes induced by ssDNA were probed in our experiments and depicted in the modelled structures.

Since it has been shown that the χ subunit binds to the flexible C-terminal tail of SSB through as a long, unconserved linker (green spheres in Fig. 4),[9,11] connectivity between an SSB subunit and χ is enforced. We therefore generate structural models of ψχ–SSB_4 by sampling the full range of conformational space whilst maintaining the SSB subunit and χ interaction. After filtering 50,000 models of ψχ–SSB_4 using the overlap restraint, a total of 22,569 structures remained for further analysis. These structures are subjected to scoring, using volume and CCS restraints. The structures receiving the top 1% of scores revealed a major cluster that represents ~95% of the best-scoring models. A representative structure for ψχ–SSB_4, selected as the best-scored model within the cluster, is shown for the hexameric subcomplex (Fig. 4a, top right).

To investigate the effect of ssDNA on the ψχ–SSB_4 architecture, we compare the CCS for SSB_4–ssDNA and SSB_4. The CCS of the DNA bound form was found to be ~10% larger than that for SSB_4 (Table 1). Moreover, the X-ray crystal structure of the structured portion of SSB_4 bound to ssDNA revealed that ssDNA wraps extensively around the tetramer; therefore, its contribution to the overall CCS is expected to be minor and within the anticipated error of the CCS restraint (overall ~5.5% total increase in CCS when DNA is bound) (PDB ID: 1EYG[42]) (Fig. S2). Such a structure is known to undergo a conformational change within the SSB_4–ssDNA complex. To perform modelling on this subassembly, we built a low-resolution model for SSB_4–ssDNA, based on the available X-ray crystal structure of SSB_4 (Fig. S2).

The next step was to build the ψχ–SSB_4–ssDNA subcomplex from its constituent parts. After applying the overlap restraint (described above) and by enforcing connectivity between SSB and the χ subunit, a total of 26,987 structures remained for further analysis. These models were subsequently evaluated and the 1% best-scoring models revealed two major clusters, 79% and 21%, respectively (Fig. 4b). The best-scored model of the major cluster suggests a rather extended arrangement and represents the most likely architecture of ψχ–SSB_4–ssDNA.

Structural characterisation of the full clamp loader complex bound to SSB_4

We turned our attention to the interactions between the full clamp loader and SSB_4 with and without DNA bound. MS analysis of the $\gamma_3\delta\delta'\psi\chi$–$SSB_4$ containing solution revealed the presence of the $\gamma_3\delta\delta'\psi\chi$ complex and a low abundance of peaks corresponding to the intact $\gamma_3\delta\delta'\psi\chi$–$SSB_4$ assembly (Fig. 5a). To investigate the effect of ssDNA binding, we also carried out MS on $\gamma_3\delta\delta'\psi\chi$–$SSB_4$ in the presence of ssDNA (Fig. 5b). From the MS spectrum, three charge state series corresponding to SSB_4, $\gamma_3\delta\delta'\psi\chi$, and $\gamma_3\delta\delta'\psi\chi$–$SSB_4$–$ssDNA_2$ were observed (Fig. 5b). Similar to ψχ–SSB_4 and $\gamma_3\delta\delta'\psi\chi$–$SSB_4$, the peak intensity of ions for $\gamma_3\delta\delta'\psi\chi$–$SSB_4$–$ssDNA_2$ was low, highlighting the weak interactions between ψχ and SSB_4. However, we observed increase in peak intensity of $\gamma_3\delta\delta'\psi\chi$–$SSB_4$–$dC_{35}$, suggesting enhanced stability of $\gamma_3\delta\delta'\psi\chi$–$SSB_4$ complex when ssDNA is bound. Interestingly and contrary to ψχ–SSB_4 subcomplex formation (Fig. 4b), two ssDNA (d35-mer) were found bound to SSB_4 using MS, consistent with the X-ray crystal structure (Table S1).

To quantify any structural differences between the $\gamma_3\delta\delta'\psi\chi$–$SSB_4$ in its DNA bound or unbound form, we examined the IM ATD for the two species (Fig. 5d). Measured drift times were converted into CCSs. The CCS for $\gamma_3\delta\delta'\psi\chi$–$SSB_4$ was measured as 11,529 ± 345 $Å^2$, almost 10% lower than that of the CCS measured when ssDNA is present (12,488 $Å^2$) (Table 1). The experimental error of IM measurements is estimated to be ~3% (2 SDs from the mean is used to define the error in the CCS restraint, i.e., 6%).[46] X-ray crystallography has shown that ssDNA wraps around SSB_4 by simultaneously inducing a conformational change in the protein.[9,42] Therefore, as binding of ssDNA does not significantly increase the CCS of SSB_4 (Fig. S2), we attribute the difference in our experiments to conformational change in SSB_4 upon binding to ssDNA.

To build a three-dimensional model for the eleven-subunit complex, we used the best-scored structures for $\gamma_3\delta\delta'\psi\chi$ from the previous steps (Figs. 3 and 4) and model structures of SSB_4 and SSB_4–ssDNA (Fig. S2) as input for sampling with the Monte Carlo search. The sampled structures were restrained by enforcing connectivity between the χ subunit and C-terminal tail of SSB_4 as suggested previously.[9,47] As mentioned above, after initial filtering using the overlap restraint, 21,902 and 20,189 models remained for $\gamma_3\delta\delta'\psi\chi$–$SSB_4$ and $\gamma_3\delta\delta'\psi\chi$–$SSB_4$–ssDNA2, respectively. These models were subsequently evaluated using the scoring function obtained from CCS and volume restraints. Next, we clustered the best-scored models revealing two main clusters for the structure with the presence of ssDNA and three for that without ssDNA. The representative structures of the major clusters for $\gamma_3\delta\delta'\psi\chi$–$SSB_4$ (56.2%) and $\gamma_3\delta\delta'\psi\chi$–$SSB_4$–$ssDNA_2$ (46.8%) were selected as the final solutions (Fig. 5c and d and Fig. S2). The representative structures for $\gamma_3\delta\delta'\psi\chi$–$SSB_4$ with and without ssDNA suggest elongated arrangements (Fig. 5c and e).

Discussion

Here, we used restraints derived from IM-MS to assemble an important multiprotein complex belong-

ing to the AAA + class of motors, the seven-subunit *E. coli* DNA clamp loader complex together with the SSB homotetramer. To the best of our knowledge, this is the largest heteromeric complex that has been assembled from its building blocks primarily using information from IM-MS.

The first crucial step in our approach was to define the stoichiometry of subcomplexes and their subunit connectivity from MS experiments and other available information in the literature. This information is translated into spatial restraints and, together with CCSs derived from IM, is integrated into our modelling approach for generating candidate model structures. Four different restraints are used: subcomplex connectivity, overlap between subunits and domains, volume, and CCS. The connectivity restraint is used to generate a topological map of subunits and to restrain the model building process thus leading to smaller sampling space. Overlap and volume restraints ensure realistic physical packing between interacting subunits and domains. The CCS restraint provides information on the overall shape of complexes and its use relies on the assumption that subcomplexes observed in the gas phase have not undergone significant structural rearrangement. Although using these restraints independently is, in most cases, not sufficient to predict the native structures, the integration of all four restraints into a scoring function and subsequent analysis can uncover structural features of otherwise intractable assemblies.

The most important step in our modelling strategy is to translate accurately all available information into a single scoring function. After scoring all candidate models, we assessed their structural variability of the best-scored models by performing clustering analysis (see Supplementary Data). The two clusters identified for the seven-subunit clamp loader complex yielded no unique solution for this complex, although over 80% of models fall in the major cluster. It is interesting to note that the two conformations of clamp loader complex found here could possibly describe an ensemble of highly dynamic structures consistent with nature of this complex reported previously. Similar results were obtained for the $\psi\chi$–SSB_4 subcomplex in the presence of ssDNA, whereas for the $\psi\chi$–SSB_4 without ssDNA, only one cluster was identified. This leads to the conclusion that the proposed structure for $\psi\chi$–SSB_4 is the only solution given the input data. The fact that we observe only one cluster for $\psi\chi$–SSB_4 complex and two clusters for $\psi\chi$–SSB_4–ssDNA is attributed to the symmetry in the structure of SSB_4 in the absence of ssDNA. Finally, for the clamp loader–SSB_4 complex, two and three major clusters are identified within the best-scored models for the complex without and with ssDNA, respectively (Fig. S3). Despite the fact that we selected the major clusters (~56% and 47%) as a solution for two conformations of clamp loader–SSB_4, with and without the ssDNA, the probability that the representative structure of the minor clusters (~40% for $\gamma_3\delta\delta'$–$\psi\chi$–SSB_4 and ~26% for $\gamma_3\delta\delta'$–$\psi\chi$–SSB_4–$ssDNA_2$) could also be a solution cannot be neglected. This ambiguity could be attributed to the possibility that both conformations most likely populate multiple states in solution and/or in the gas phase. Possibly, the dynamic nature of those complexes is further enhanced by the highly flexible regions identified in some of their constituent subunits.

It has been proposed that SSB protein interacts with $\psi\chi$ during replication in a way that stabilises the complex by bringing together essential entities for clamp loading and replication to occur.[12] Moreover, a critical aspect for rapid DNA replication is the interaction of ssDNA with the SSB protein.[15] Both $\psi\chi$–SSB_4 and $\psi\chi$–SSB_4–ssDNA subcomplexes were observed in our experiments, accommodating different conformations, contrary to a previous study suggesting that $\psi\chi$–SSB interaction occurs only when SSB is bound to the ssDNA.[12] Therefore, based on our data, we suggest that the presence of DNA is essential for subcomplex activity but does not potentiate binding of $\psi\chi$ to SSB. Such binding is potentiated by the presence of the χ subunit, which also has the functional role in promoting sliding clamp assembly.[45]

The $\gamma_3\delta\delta'$ complex within the Pol III holoenzyme interacts with SSB_4 through the $\psi\chi$. Such an interaction, which is maintained during the transition between initiation and elongation, is known to increase the efficiency of the latter and to aid the clamp loading activity.[45] By measuring CCSs, we observed structural differences between the bound and unbound forms of the $\gamma_3\delta\delta'$–$\psi\chi$–SSB_4 complex, which we attribute to the conformational changes induced by ssDNA binding. This conformational flexibility is in line with the suggested theory of a DNA-triggered switch between active and inactive states upon binding of ssDNA. It is interesting to speculate that the larger conformation proposed for the active state of $\gamma_3\delta\delta'$–$\psi\chi$–SSB_4 is adopted to aid clamp loading activity onto DNA.

Overall in this study, we first established the stoichiometry of subunit and ssDNA interactions in the $\gamma_3\delta\delta'$–$\psi\chi$–SSB_4 complex wherein two copies of ssDNA were found to be present, as opposed to $\psi\chi$–SSB_4 and SSB_4 where only one bound ssDNA was observed. By applying an integrative modelling approach, which integrates structural information from IM-MS, we propose three-dimensional structural models of the eleven-subunit $\gamma_3\delta\delta'$–$\psi\chi$–SSB_4 complex and its subcomplexes observed in our MS experiments. Overall, our study highlights the power of IM-MS when combined with integrative modelling strategies to probe structural and dynamical aspects of heteromeric complexes in advance of other structural information from classical techniques.

Materials and Methods

Experimental procedure

Individual subunits (γ, δ, δ′, ψχ, and SSB) were overexpressed in *E. coli* and purified as described elsewhere.[10,40,48–50] For MS of modules, χψ, $\gamma_3\delta\delta'$, and $\gamma_3\delta\delta'\psi\chi$, aliquots were buffer exchanged into 0.1 M NH_4OAc at pH 7.6 by using Vivaspin 500 concentrators (Sartorius, UK) with a 10- to 50-kDa molecular mass cutoff, depending on the size of the complex. SSB was dialysed in 1 M NH_4OAc using a Slide-A-Lyzer Mini Dialysis unit (Pierce, UK) with a 10-kDa molecular mass cutoff. Concentrations for χψ, $\gamma_3\delta\delta'$, $\gamma_3\delta\delta'\psi\chi$, and SSB_4 were measured spectrophotometrically at 280 nm, using ε_{280} = 53,680, 170,090, 223,770, and 111,520 M^{-1} cm^{-1}, respectively. The final concentrations of the proteins were in a range 5–10 μM. Two ssDNA oligomers (polyC 35-mer and 65-mer) were synthesised (Integrated DNA Technologies, Belgium). For MS, the oligomers were diluted to 10 μM in 0.5 M NH_4OAc at pH 6.9. In the case of $\gamma_3\delta\delta'\psi\chi$–$SSB_4$–ssDNA complex formation, SSB_4, $\gamma_3\delta\delta'\psi\chi$, and ssDNA were incubated at equimolar ratios of 1:1:1 after buffer exchange.

Ion mobility mass spectrometry

IM-MS measurements were carried out on a modified Synapt HDMS system (Waters Corporation, UK) described in detail elsewhere.[46] Briefly, the travelling-wave IM cell was replaced with a linear RF-confining drift tube, allowing direct CCS measurements. Drift times were calculated from the well-described relationship for an ion travelling through a gas in the presence of a weak electric field.[51–53] Instrumental parameters were typically as follows: capillary voltage, 1.4 kV; cone voltage, 40 V; trap collision energy, 12 V; backing pressure, 6 mbar. The IMS (*i*on *m*obility *s*pectrometry) cell contained helium gas at a pressure of 2–2.5 Torr. Drift times were typically measured at 10 drift voltages ranging from 50 to 200 V.

Mass spectra for SSB_4 binding to ssDNA were acquired using a modified Q-TOF2 mass spectrometer (Waters Corporation, UK).[54] MS experiments were conducted with capillary voltages up to 1.8 kV, a sample cone of 100 V, an extractor voltage of 10 V, an ion transfer stage pressure of 4.0×10^{-3} mbar, a quadruple analyser pressure of 1.8×10^{-5} mbar, and a time-of-flight analyser pressure of 4.7×10^{-7} mbar.

Calculated theoretical CCSs

The CCSs were calculated using MOBCAL[55] and a version adapted in-house for CG models.[56] Two representation approaches (all-atom and CG) were employed in this study. The projection approximation (PA) method[57] was employed for calculating CCS. Briefly, this method calculates numerically the orientationally averaged CCS of the protein. Although the PA method is known to underestimate the experimental CCS by neglecting multiple collisions between the ion and buffer gas,[58,59] it is the only method capable of calculating CCS for CG models based on spheres.[36] Furthermore, it has been shown to correlate well with experimental CCS for globular protein complexes ($R^2 > 0.99$).[60] We therefore made use of a scaled PA CCS, whereby the experimental CCS can be predicted (±3%) by multiplying the PA CCS by a factor of 1.14.[60]

Acknowledgements

We thank Drs Slobodan Jergic and Claire Mason (University of Wollongong, Australia) for highly purified, homogeneous proteins and meticulously isolated multiprotein complexes. We acknowledge funding from the Biotechnology and Biological Sciences Research Council (to A.P. and A.Y.P.), the Royal Society (to C.V.R.), and Waters (to B.T.R.). Julien Marcoux (University of Oxford) is acknowledged for critical reading of the manuscript.

Supplementary Data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2013.04.006

Received 25 January 2013;
Received in revised form 14 March 2013;
Accepted 3 April 2013
Available online 11 April 2013

Keywords:
integrative modelling;
ion mobility-mass spectrometry;
single-stranded DNA binding protein (SSB);
clamp loader;
DNA replication

Present address: A. Y. Park, Institut für Pharmazeutische Chemie, Philipps-Universität Marburg, Marbacher Weg 6, 35032 Marburg, Germany.

† A.P. and A.Y.P. contributed equally to this work.

Abbreviations used:
IM, ion mobility; ssDNA, single-stranded DNA; MS, mass spectrometry; CCS, collision cross-section; CG, coarse grained; ATD, arrival time distribution; PDB, Protein Data Bank; USR, ultrafast shape recognition; PA, projection approximation.

References

1. Kornberg, A. & Baker, T. A. (1992). *DNA Replication*, 2nd Ed. W. H, Freeman, New York.

2. Schaeffer, P. M., Headlam, M. J. & Dixon, N. E. (2005). Protein–protein interactions in the eubacterial replisome. *IUBMB Life*, **57**, 5–12.
3. Johnson, A. & O'Donnell, M. (2005). Cellular DNA replicases: components and dynamics at the replication fork. *Annu. Rev. Biochem.* **74**, 283–315.
4. Kelch, B. A., Makino, D. L., O'Donnell, M. & Kuriyan, J. (2011). How a DNA polymerase clamp loader opens a sliding clamp. *Science*, **334**, 1675–1680.
5. Pomerantz, R. T. & O'Donnell, M. (2007). Replisome mechanics: insights into a twin DNA polymerase machine. *Trends Microbiol.* **15**, 156–164.
6. Benkovic, S. J., Valentine, A. M. & Salinas, F. (2001). Replisome-mediated DNA replication. *Annu. Rev. Biochem.* **70**, 181–208.
7. Gulbis, J. M., Kazmirski, S. L., Finkelstein, J., Kelman, Z., O'Donnell, M. & Kuriyan, J. (2004). Crystal structure of the chi:psi sub-assembly of the *Escherichia coli* DNA polymerase clamp-loader complex. *Eur. J. Biochem.* **271**, 439–449.
8. Tsuchihashi, Z. & Kornberg, A. (1990). Translational frameshifting generates the gamma subunit of DNA polymerase III holoenzyme. *Proc. NAtl Acad. Sci. USA*, **87**, 2516–2520.
9. Simonetta, K. R., Kazmirski, S. L., Goedken, E. R., Cantor, A. J., Kelch, B. A., McNally, R. *et al.* (2009). The mechanism of ATP-dependent primer-template recognition by a clamp loader complex. *Cell*, **137**, 659–671.
10. Wijffels, G., Dalrymple, B. P., Prosselkov, P., Kongsuwan, K., Epa, V. C., Lilley, P. E. *et al.* (2004). Inhibition of protein interactions with the beta 2 sliding clamp of Escherichia coli DNA polymerase III by peptides from beta 2-binding proteins. *Biochemistry*, **43**, 5661–5671.
11. Yuzhakov, A., Kelman, Z. & O'Donnell, M. (1999). Trading places on DNA–a three-point switch underlies primer handoff from primase to the replicative DNA polymerase. *Cell*, **96**, 153–163.
12. Glover, B. P. & McHenry, C. S. (1998). The chi psi subunits of DNA polymerase III holoenzyme bind to single-stranded DNA-binding protein (SSB) and facilitate replication of an SSB-coated template. *J. Biol. Chem.* **273**, 23476–23484.
13. Chase, J. W. & Williams, K. R. (1986). Single-stranded DNA binding proteins required for DNA replication. *Annu. Rev. Biochem.* **55**, 103–136.
14. Meyer, R. R. & Laine, P. S. (1990). The single-stranded DNA-binding protein of Escherichia coli. *Microbiol. Rev.* **54**, 342–380.
15. Shereda, R. D., Kozlov, A. G., Lohman, T. M., Cox, M. M. & Keck, J. L. (2008). SSB as an organizer/mobilizer of genome maintenance complexes. *Crit. Rev. Biochem. Mol. Biol.* **43**, 289–318.
16. Ason, B., Bertram, J. G., Hingorani, M. M., Beechem, J. M., O'Donnell, M., Goodman, M. F. & Bloom, L. B. (2000). A model for Escherichia coli DNA polymerase III holoenzyme assembly at primer/template ends. DNA triggers a change in binding specificity of the gamma complex clamp loader. *J. Biol. Chem.* **275**, 3006–3015.
17. Marceau, A. H., Bahng, S., Massoni, S. C., George, N. P., Sandler, S. J., Marians, K. J. & Keck, J. L. (2011). Structure of the SSB-DNA polymerase III interface and its role in DNA replication. *EMBO J.* **30**, 4236–4247.
18. Jeruzalmi, D., O'Donnell, M. & Kuriyan, J. (2001). Crystal structure of the processivity clamp loader gamma (gamma) complex of E. coli DNA polymerase III. *Cell*, **106**, 429–441.
19. Hernandez, H., Dziembowski, A., Taverner, T., Seraphin, B. & Robinson, C. V. (2006). Subunit architecture of multimeric complexes isolated directly from cells. *EMBO Rep.* **7**, 605–610.
20. Zhou, M., Sandercock, A. M., Fraser, C. S., Ridlova, G., Stephens, E., Schenauer, M. R. *et al.* (2008). Special Feature: Mass spectrometry reveals modularity and a complete subunit interaction map of the eukaryotic translation factor eIF3. *Proc. Natl Acad. Sci. USA.*
21. Sharon, M., Mao, H., Boeri Erba, E., Stephens, E., Zheng, N. & Robinson, C. V. (2009). Symmetrical modularity of the COP9 signalosome complex suggests its multifunctionality. *Structure*, **17**, 31–40.
22. Taverner, T., Hernandez, H., Sharon, M., Ruotolo, B. T., Matak-Vinkovic, D., Devos, D. *et al.* (2008). Subunit architecture of intact protein complexes from mass spectrometry and homology modeling. *Acc. Chem. Res.* **41**, 617–627.
23. Zhou, M., Morgner, N., Barrera, N. P., Politis, A., Isaacson, S. C., Matak-Vinkovic, D. *et al.* (2011). Mass spectrometry of intact V-type ATPases reveals bound lipids and the effects of nucleotide binding. *Science*, **334**, 380–385.
24. Loo, J. A., Berhane, B., Kaddis, C. S., Wooding, K. M., Xie, Y., Kaufman, S. L. & Chernushevich, I. V. (2005). Electrospray ionization mass spectrometry and ion mobility analysis of the 20S proteasome complex. *J. Am. Soc. Mass Spectrom.* **16**, 998–1008.
25. Ruotolo, B. T., Giles, K., Campuzano, I., Sandercock, A. M., Bateman, R. H. & Robinson, C. V. (2005). Evidence for macromolecular protein rings in the absence of bulk water. *Science*, **310**, 1658–1661.
26. Robinson, C. V., Sali, A. & Baumeister, W. (2007). The molecular sociology of the cell. *Nature*, **450**, 973–982.
27. Smith, D. P., Radford, S. E. & Ashcroft, A. E. (2010). Elongated oligomers in beta2-microglobulin amyloid assembly revealed by ion mobility spectrometry-mass spectrometry. *Proc. NAtl Acad. Sci. USA*, **107**, 6794–6798.
28. Bernstein, S. L., Dupuis, N. F., Lazo, N. D., Wyttenbach, T., Condron, M. M., Bitan, G. *et al.* (2009). Amyloid-b protein oligomerization and the importance of tetramers and dodecamers in the aetiology of alzheimer's disease. *Nat. Chem.* **1**, 326–331.
29. Grabenauer, M., Wyttenbach, T., Sanghera, N., Slade, S. E., Pinheiro, T. J., Scrivens, J. H. & Bowers, M. T. (2010). Conformational stability of Syrian hamster prion protein PrP(90–231). *J. Am. Chem. Soc.* **132**, 8816–8818.
30. Hilton, G. R., Lioe, H., Stengel, F., Baldwin, A. J. & Benesch, J. L. (2012). Small heat-shock proteins: paramedics of the cell. *Top. Curr. Chem.*
31. Pukala, T. L., Ruotolo, B. T., Zhou, M., Politis, A., Stefanescu, R., Leary, J. A. & Robinson, C. V. (2009). Subunit architecture of multiprotein assemblies determined using restraints from gas-phase measurements. *Structure*, **17**, 1235–1243.
32. Wang, S. C., Politis, A., Di Bartolo, N., Bavro, V. N., Tucker, S. J., Booth, P. J. *et al.* (2010). Ion mobility

mass spectrometry of two tetrameric membrane protein complexes reveals compact structures and differences in stability and packing. *J. Am. Chem. Soc.* **132**, 15468–15470.
33. van Duijn, E., Barbu, I. M., Barendregt, A., Jore, M. M., Wiedenheft, B., Lundgren, M. *et al.* (2012). Native tandem and ion mobility mass spectrometry highlight structural and modular similarities in CRISPR-associated protein complexes from Escherichia coli and Pseudomonas aeruginosa. *Mol. Cell. Proteomics.*
34. Uetrecht, C., Barbu, I. M., Shoemaker, G. K., van Duijn, E. & Heck, A. J. (2011). Interrogating viral capsid assembly with ion mobility-mass spectrometry. *Nat. Chem.* **3**, 126–132.
35. Heck, A. J. (2008). Native mass spectrometry: a bridge between interactomics and structural biology. *Nat. Methods*, **5**, 927–933.
36. Hall, Z., Politis, A. & Robinson, C. V. (2012). Structural modeling of heteromeric protein complexes from disassembly pathways and ion mobility-mass spectrometry. *Structure*, **20**, 1596–1609.
37. Politis, A., Park, A. Y., Hyung, S. J., Barsky, D., Ruotolo, B. T. & Robinson, C. V. (2010). Integrating ion mobility mass spectrometry with molecular modelling to determine the architecture of multiprotein complexes. *PLoS One*, **5**, e12080.
38. Russel, D., Lasker, K., Webb, B., Velazquez-Muriel, J., Tjioe, E., Schneidman-Duhovny, D. *et al.* (2012). Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol.* **10**, e1001244.
39. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E. *et al.* (2005). Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802.
40. Park, A. Y., Jergic, S., Politis, A., Ruotolo, B. T., Hirshberg, D., Jessop, L. L. *et al.* (2010). A single subunit directs the assembly of the Escherichia coli DNA sliding clamp loader. *Structure*, **18**, 285–292.
41. Matsumoto, T., Morimoto, Y., Shibata, N., Kinebuchi, T., Shimamoto, N., Tsukihara, T. & Yasuoka, N. (2000). Roles of functional loops and the C-terminal segment of a single-stranded DNA binding protein elucidated by X-Ray structure analysis. *J. Biochem.* **127**, 329–335.
42. Raghunathan, S., Kozlov, A. G., Lohman, T. M. & Waksman, G. (2000). Structure of the DNA binding domain of E. coli SSB bound to ssDNA. *Nat. Struct. Biol.* **7**, 648–652.
43. Johnson, S. (1967). Hierarchical clustering schemes. *Psychometrika*, **32**, 241–254.
44. Ballester, P. J. & Richards, W. G. (2007). Ultrafast shape recognition to search compound databases for similar molecular shapes. *J. Comput. Chem.* **28**, 1711–1723.
45. Kelman, Z., Yuzhakov, A., Andjelkovic, J. & O'Donnell, M. (1998). Devoted to the lagging strand-the subunit of DNA polymerase III holoenzyme contacts SSB to promote processive elongation and sliding clamp assembly. *EMBO J.* **17**, 2436–2449.
46. Bush, M. F., Hall, Z., Giles, K., Hoyes, J., Robinson, C. V. & Ruotolo, B. T. (2010). Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology. *Anal. Chem.* **82**, 9557–9565.
47. Witte, G., Urbanke, C. & Curth, U. (2003). DNA polymerase III chi subunit ties single-stranded DNA binding protein to the bacterial replication machinery. *Nucleic Acids Res.* **31**, 4434–4440.
48. Marians, K. J. (1995). Phi X174-type primosomal proteins: purification and assay. *Methods Enzymol.* **262**, 507–521.
49. Ozawa, K., Jergic, S., Park, A. Y., Dixon, N. E. & Otting, G. (2008). The proofreading exonuclease subunit epsilon of Escherichia coli DNA polymerase III is tethered to the polymerase subunit alpha via a flexible linker. *Nucleic Acids Res.* **36**, 5074–5082.
50. Mason, C. E., Jergic, S., Lo, A. T., Wang, Y., Dixon, N. E. & Beck, J. L. (2013). Escherichia coli single-stranded DNA-binding protein: NanoESI-MS studies of salt-modulated subunit exchange and DNA binding transactions. *J. Am. Soc. Mass Spectrom.*
51. Wyttenbach, T., Kemper, P. R. & Bowers, M. T. (2001). Design of a new electrospray ion mobility mass spectrometer. *International Journal of Mass Spectrometry*, **212**, 13–23.
52. Bohrer, B. C., Mererbloom, S. I., Koeniger, S. L., Hilderbrand, A. E. & Clemmer, D. E. (2008). Biomolecule analysis by Ion mobility spectrometry. *Annu. Rev. Anal. Chem.* **1**, 293–327.
53. Kanu, A. B., Gribb, M. M. & Hill, H. H. (2008). Predicting optimal resolving power for ambient pressure ion mobility spectrometry. *Anal. Chem.* **80**, 6610–6619.
54. Sobott, F., Hernandez, H., McCammon, M. G., Tito, M. A. & Robinson, C. V. (2002). A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal. Chem.* **74**, 1402–1407.
55. Mesleh, M. F., Hunter, J. M., Shavartsburg, A. A., Schartz, G. C. & Jarrold, M. F. (1996). Structural Information from ion mobility measurements: Effects of the long-range potential. *J. Phys. Chem.* **100**, 16082–16086.
56. Ruotolo, B. T., Benesch, J. L., Sandercock, A. M., Hyung, S. J. & Robinson, C. V. (2008). Ion mobility-mass spectrometry analysis of large protein complexes. *Nat. Protoc.* **3**, 1139–1152.
57. Mack, E. (1925). Average cross-sectional areas of molecules by gaseous diffusion methods. *J. Am. Chem. Soc.*
58. Shvartsburg, A. A. & Jarrold, M. F. (1996). An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chemical Physics Letters*, **261**, 86–91.
59. Bleiholder, C., Wyttenbach, T. & Bowers, M. T. (2011). A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (I). Method. *International Journal of Mass Spectrometry*, **308**, 1–10.
60. Benesch, J. L. & Ruotolo, B. T. (2011). Mass spectrometry: come of age for structural and dynamical biology. *Curr. Opin. Struct. Biol.* **21**, 641–649.